

LISA Data Analysis

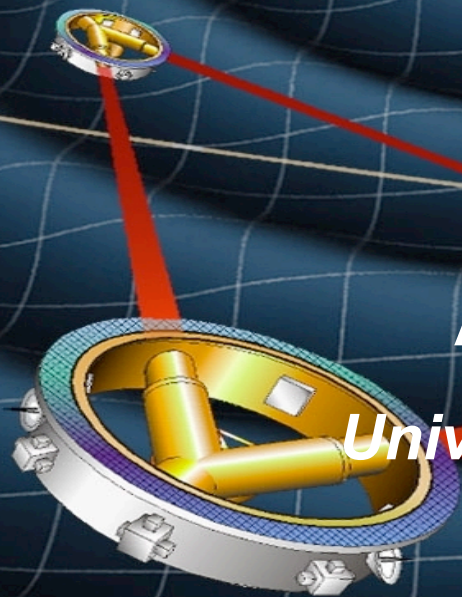
A tutorial

Alberto Vecchio

University of Birmingham (UK)

Sixth LISA Symposium

Goddard, 19th - 23rd June 2006





Outline

- The LISA data set:
 - Time delay interferometry
 - The synthesized observables
- Analysis techniques and challenges:
 - Bayesian inference: Markov Chain Monte Carlo methods
 - Matched-filtering
 - Incoherent methods
 - Hierarchical strategies
- Conclusions



LISA: a GW telescope

- LISA is an all-sky monitor:
 - All sky surveys are for free
 - “Pointing” is done in software
- LISA has guaranteed sources
- Signals are (for the vast majority) long lived
- Information about the sources are reconstructed through the structure of the recorded signal
 - Intrinsic in the waveform
 - Induced by instrument motion and response
- Each signal depends (with a few exceptions) on 7-to-17 parameters
- One year of LISA data contains:
 - Several known solar mass binaries (verification sources)
 - ~ 10000 resolvable WD binaries (a few with NS companion)
 - ~ 100 EMRIs
 - ~ 10 I/M/SMBH binaries
 - Some short lived burst events
 - Stochastic foregrounds and backgrounds



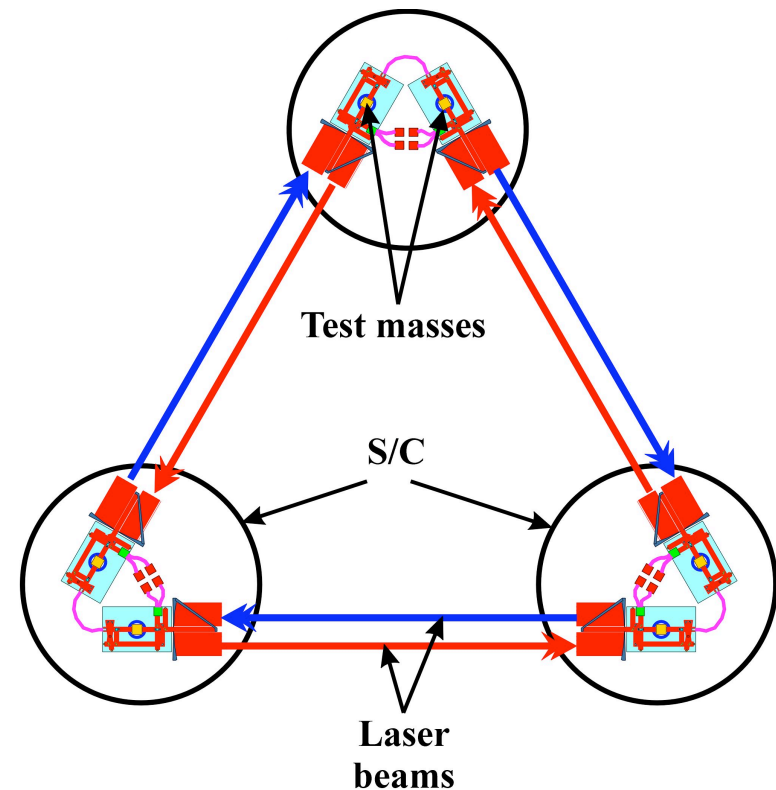
GW observations from space and ground

- **LISA**
 - Small data volume:
 $\sim 10^8$ (T/3 yr) ($f_s/1$ Hz) data points
 - Signals:
 - Many
 - Long and short lived
 - Overlapping
 - Variety of signal strength (from $h > n$ to $h \ll n$)
 - One observatory with co-located instruments
- **LIGO/GEO/VIRGO/TAMA**
 - Large data volume:
 $\sim 10^9$ (T/1 day) ($f_s/10$ kHz) data points
 - Signals:
 - Rare
 - Long and short lived
 - De facto non overlapping
 - Weak ($h \ll n$)
 - Network of several geographically separated instruments



The LISA “interferometer”

- Gravitational waves passing through the LISA constellation affect the separation between test masses
- This is monitored by comparing the locally generated frequency with the frequency of the received laser signal
- The raw data set: six 1-way Doppler links (+ housekeeping channels)



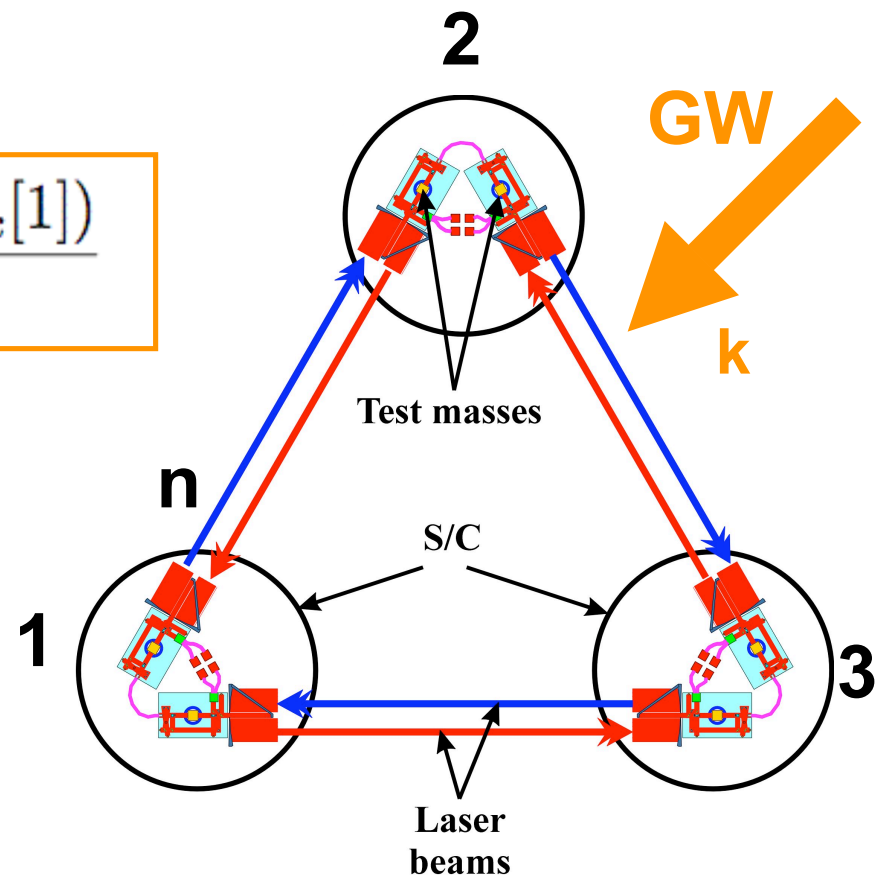


One-way Doppler link

- The effect of GWs on each one-way Doppler link is

$$y_{12}(t) = \frac{1}{2} \frac{\hat{n}^j \hat{n}^k (h_{jk}[2] - h_{jk}[1])}{1 - \hat{n}^j \hat{k}_j}$$

(Estabrook and Wahlquist 1975)





Contributions to the one-way link

- The contributions to the observable $y_{ij}(t)$ are (schematically):

$$y_{ij}(t) = C_i(t_e) - C_j(t) + n + h$$

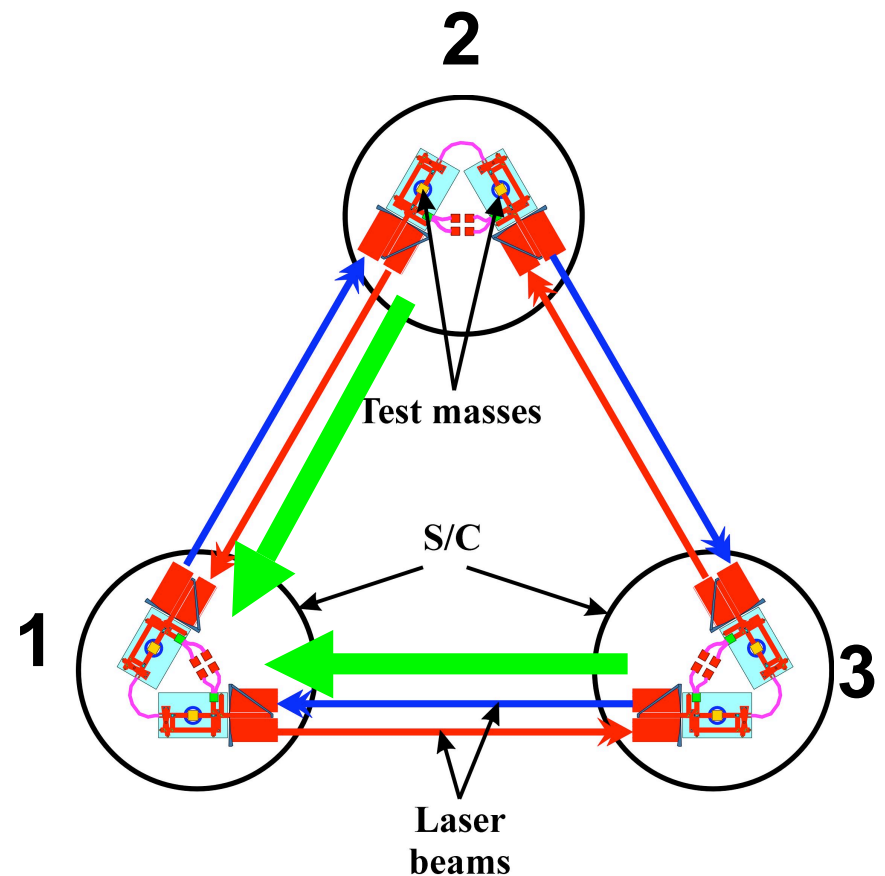
- $C(t)$: contribution due to laser frequency noise which is many orders of magnitude above h and n
 - n : secondary noises (acceleration and photon shot noise)
 - h : GWs, what we are interested in
- Key issue: how can one suppress the dominant contribution from laser noise?



Time Delay Interferometry (TDI)

- One can construct combinations that cancel e.g. $C_1(t)$:

$$y_{21}(t) - y_{31}(t) = [C_1(t) - C_2(t-L)] - [C_1(t) - C_3(t-L)]$$

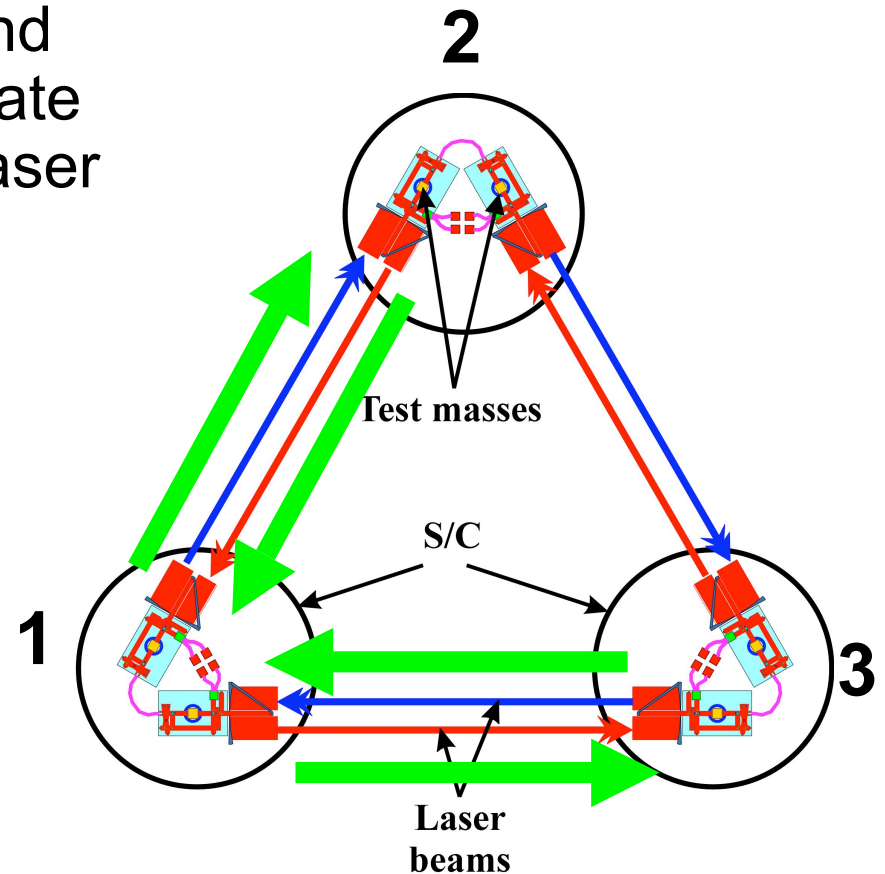




Time Delay Interferometry (TDI)

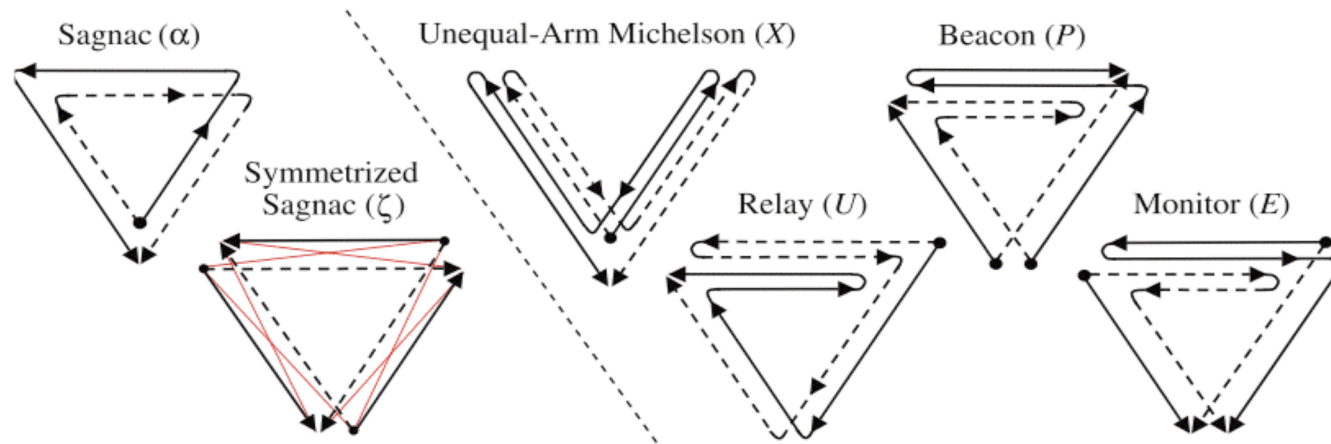
- **TDI**: linear time-delayed combinations of y_{jk} (add and subtract 1-way links to create a close loop) that cancel laser noise
- An example: **synthesized (equal-arm) Michelson interferometer**

$$\begin{aligned} [y_{21}(t) - y_{12}(t-L)] - [y_{31}(t) - y_{13}(t-L)] = \\ [C_1(t) - C_2(t-L) + C_2(t-L) - C_1(t-2L)] - \\ [C_1(t) - C_3(t-L) + C_3(t-L) - C_1(t-2L)] = \\ = \mathbf{0} + n + h \end{aligned}$$



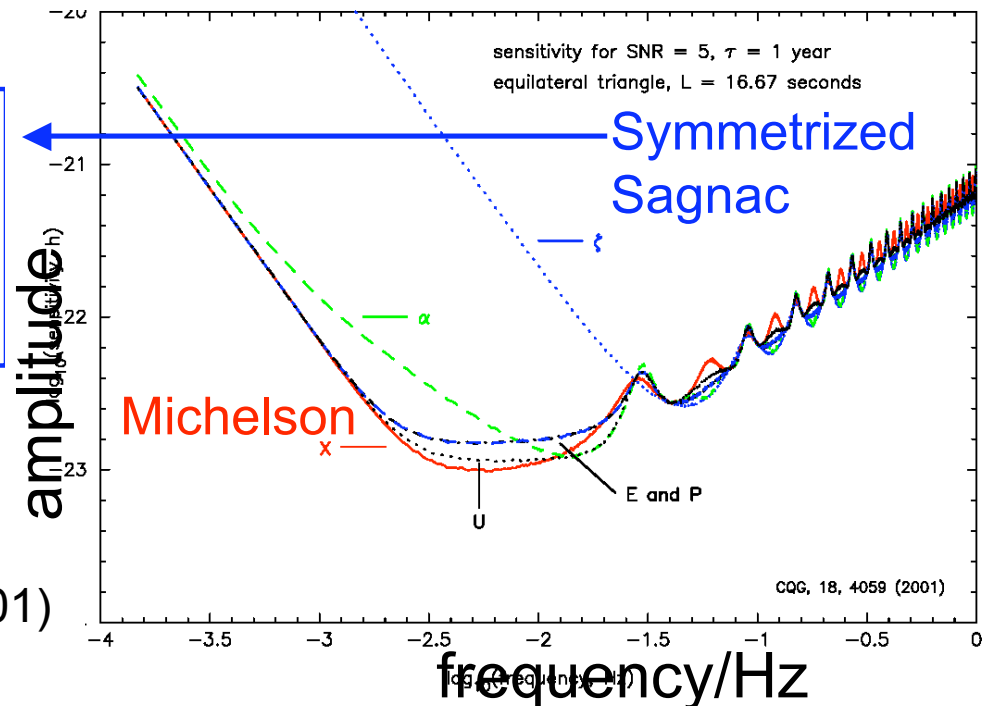


The zoo of TDI variables



Insensitive to GW at $f \ll 10$ mHz
 ($\sim 1/L$): it allows us to obtain a
noise only channel at low frequency

(Armstrong, Estabrook and Tinto, 2001)





The LISA observables: the data set

- Not all the combination are statistically independent
- Full information about the GW sky are contained in 3 independent data streams (e.g. A, E, T)
- LISA science is in $\sim 10^9$ data points
- Observables:

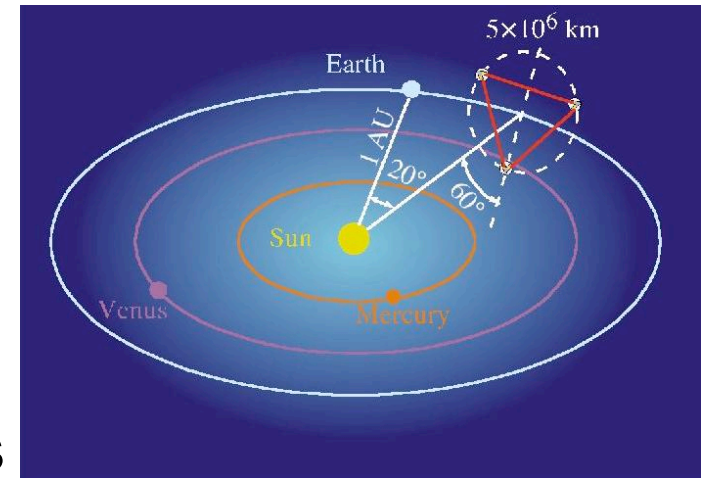
$$h(t) = \sum_k F_k(t; \text{source location}) h_k(t; \text{physics})$$

- *Many papers: Armstrong, Estabrook, Tinto, Vinet, Dhurandhar, Nayak, Vallisneri, Cornish, Larson, Prince, Shaddok, Romano, Woan.....*



Complications: TDI generations

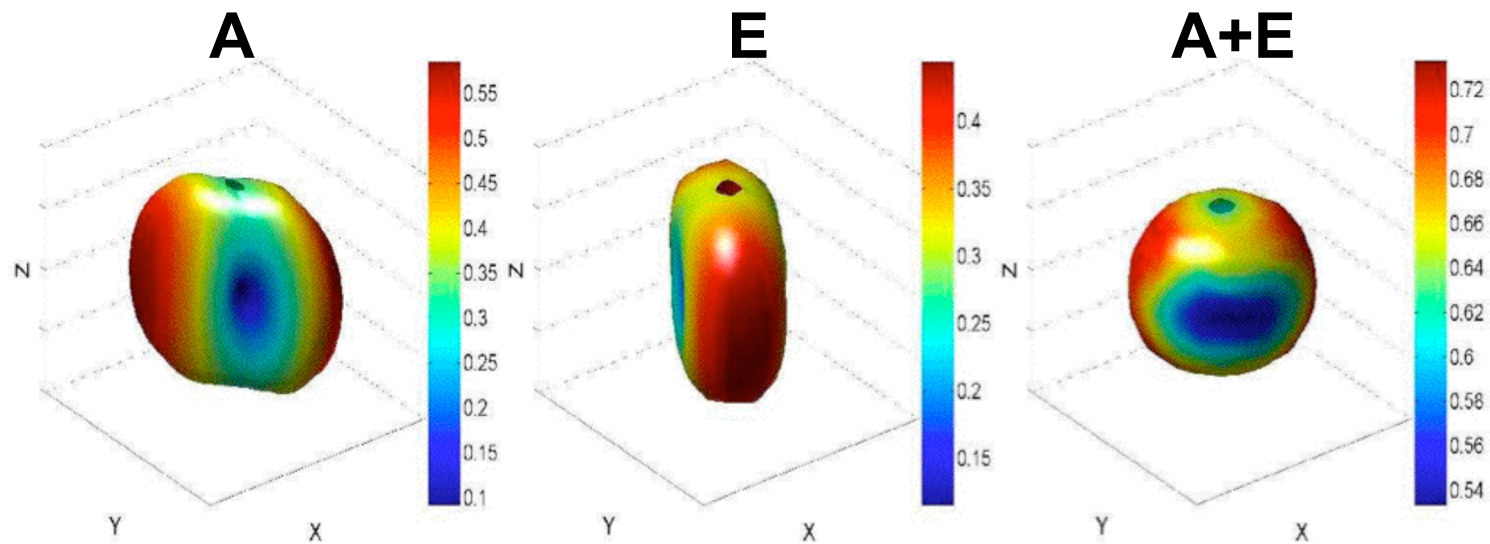
- **First generation TDI**: cancel laser noise for a static constellation
- However:
 - Constellation rotates
 - Arm length is not constant (flexing)
- **Second generation TDI**: accounts for rotation of constellation and relative motion of spacecraft
- **The second generation TDI observables are obtained as linear combinations of first generation TDI**





LISA: all-sky monitor

Sensitivity:



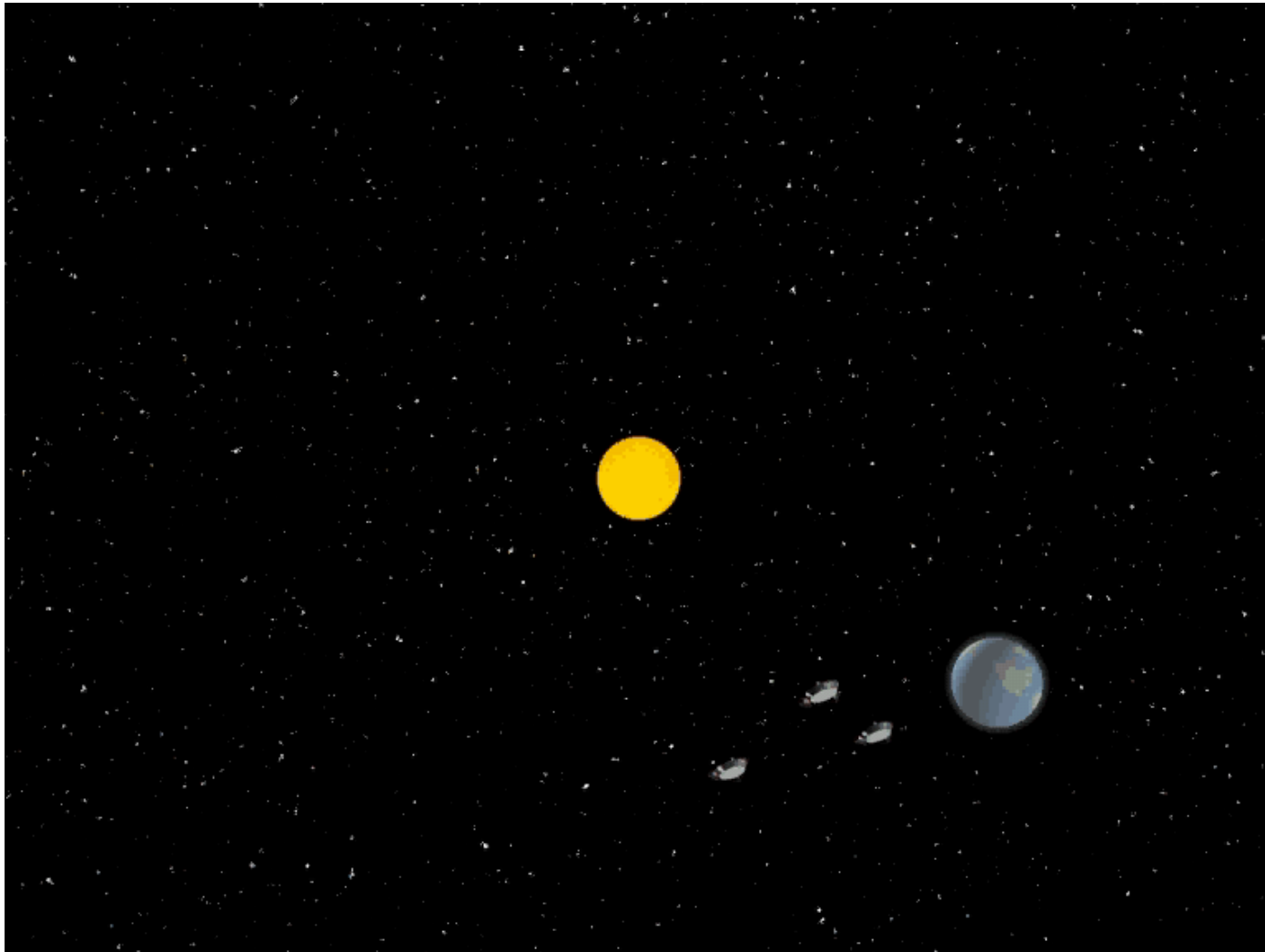
$f = 3 \text{ mHz}$

(fixed ι, ψ)

(Rogan and Bose, astro-ph/0605034)



LISA orbit



19th June 2006

A Vecchio - LISA Data Analysis

14

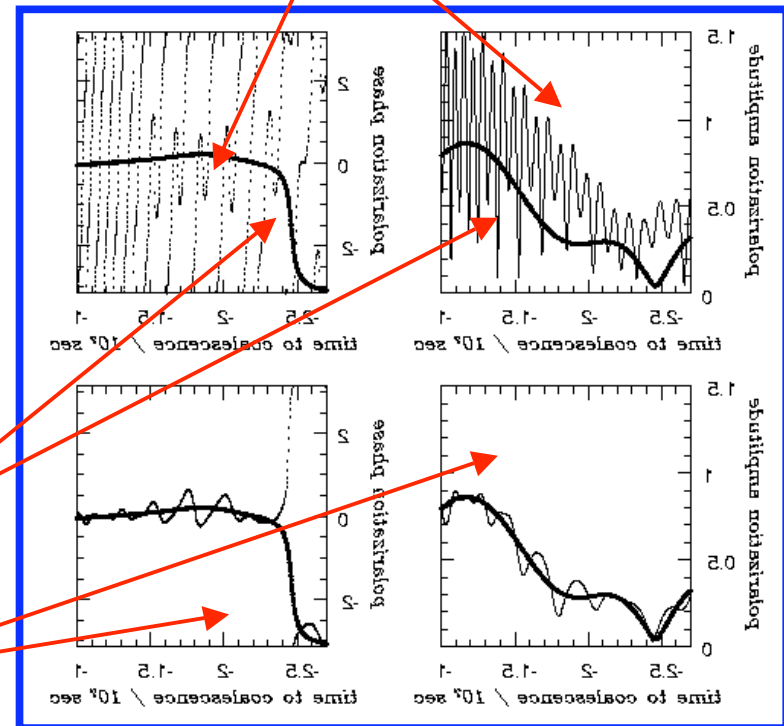
Extracting information

- Information about sources are reconstructed through structure of signal
 - Intrinsic in the waveform
 - Induced by instrument motion and response
- Example: MBH binary inspiral

LISA rotation
around its axis

Small spin

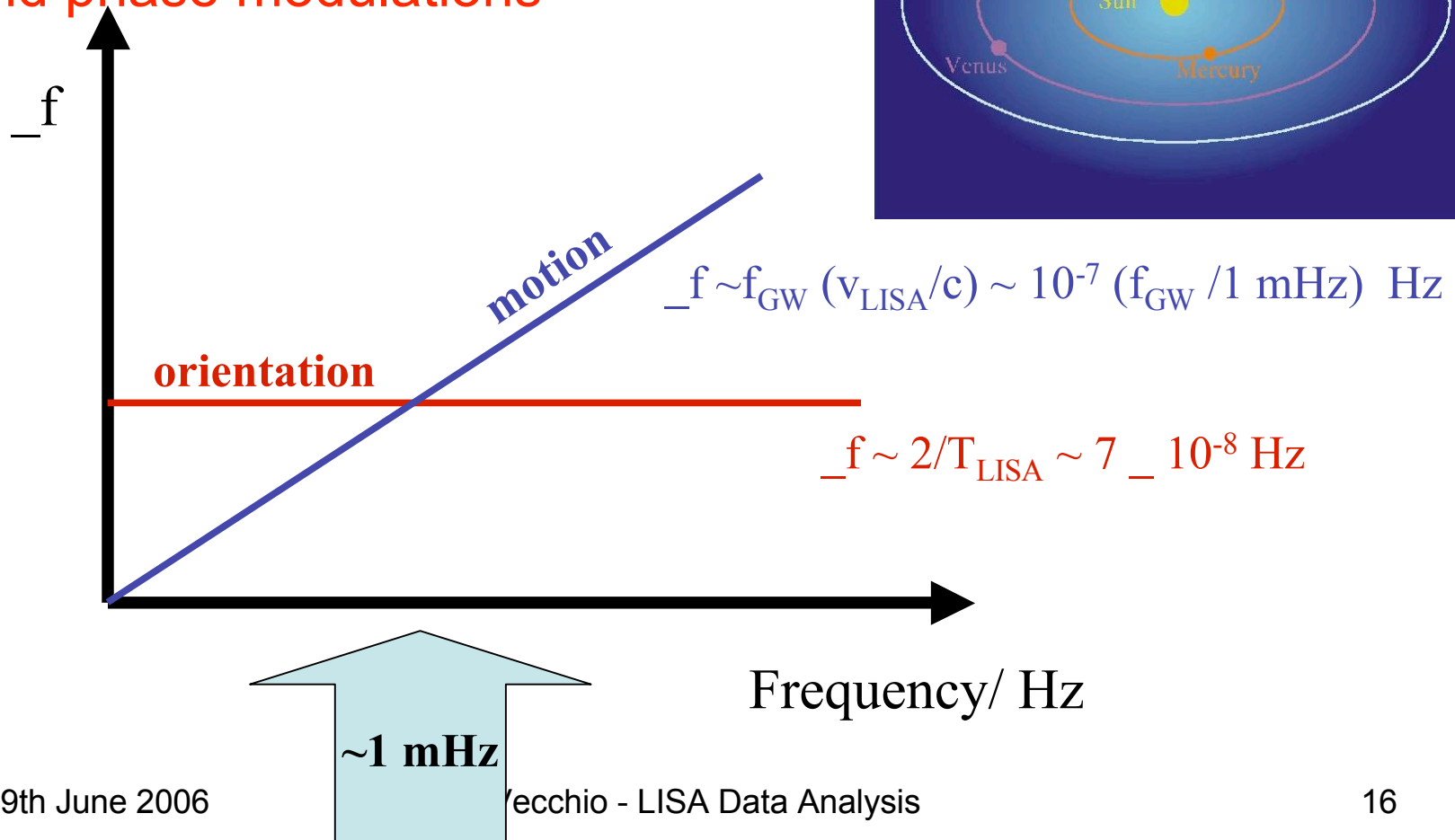
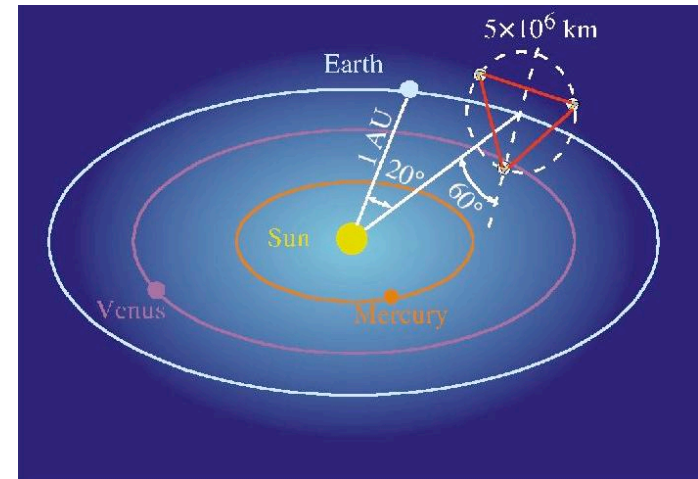
large spin





Pointing a source

Done in software: by matching
(position dependent) amplitude
and phase modulations





LISA data analysis

- Given the data set(s):

$N \gg 1$ and unknown

$$d(t) = \sum_j^N h_j(t; \vec{\lambda}_j) + n(t)$$

Waveform (convolved with the instrument response): could either be well modeled or poorly known

- We want to identify the signals and extract information on the unknown parameters λ
- Bayesian approach: derive the posterior probability density function (pdf)
- Frequentist approach: construct a detection statistic (filter the data)



LISA data analysis issues

- Source specific:
 - Identify each source down to the detection threshold
 - Detect Extreme Mass-Ratio Inspirals (EMRIs)
 - Unambiguously detect a stochastic signals
 -
 - Many similarities with LIGO/GEO/VIRGO
- Global analysis:
 - Extract information (do astronomy, cosmology and fundamental physics) with large number of overlapping sources (both loud and weak signals)
 - ... so large they become confused in a significant portion of the observational window



Bayesian inference

- Bayes' theorem: the appropriate rule for updating our degree of belief (in one of several hypotheses within some world view I) when we have new data:

$$p(\text{model}|\text{data}, I) = \frac{p(\text{data}|\text{model}, I) p(\text{model}|I)}{p(\text{data}|I)}$$

Diagram illustrating Bayes' theorem with labels and arrows:

- Posterior** (red text) points to $p(\text{model}|\text{data}, I)$
- Likelihood** (red text) points to $p(\text{data}|\text{model}, I)$
- Prior** (red text) points to $p(\text{model}|I)$
- Evidence, or "global likelihood"** (red text) points to $p(\text{data}|I)$

- A consequence of the product rule:

$$p(a|b) p(b) = p(b|a) p(a)$$



Technical problem: integration

- The model - $\sum_j h_j(\lambda)$ - depends on a very large number of parameters ($\sim 10^5$)
- We are usually interested in pdf's of one parameter at the time: marginalization

$$p(\lambda_j) = \int \dots \int p(\vec{\lambda}) d\lambda_1 d\lambda_{j-1} \dots d\lambda_{j+1} d\lambda_N$$

- The difficulty is the integration (large number of dimensions)



Markov Chain Monte Carlo (MCMC) methods

- We need to evaluate integrals of the form:

$$p(\lambda_j) = \int \dots \int p(\vec{\lambda}) d\lambda_1 d\lambda_{j-1} \dots d\lambda_{j+1} d\lambda_N$$

- The strategy is to sample the space $(\lambda_1, \lambda_2, \dots, \lambda_N)$ so that the density of the sample reflects the posterior probability $p(\lambda_1, \lambda_2, \dots, \lambda_N)$
- MCMC algorithms perform random walks in the parameter space so that the probability of being in a hypervolume dV is $p dV$
- The random walk is a Markov chain: the transition probability of making a step depends on the proposed location $x' (\lambda_1, \lambda_2, \dots, \lambda_N)$ *and* the current location $x(\lambda_1, \lambda_2, \dots, \lambda_N)$
- MCMC methods have demonstrated success in problems with large parameter number (Google, financial markets, WMAP....)



An algorithm: Metropolis-Hastings

We want to derive $p(x)$

Assume we are at location x_t

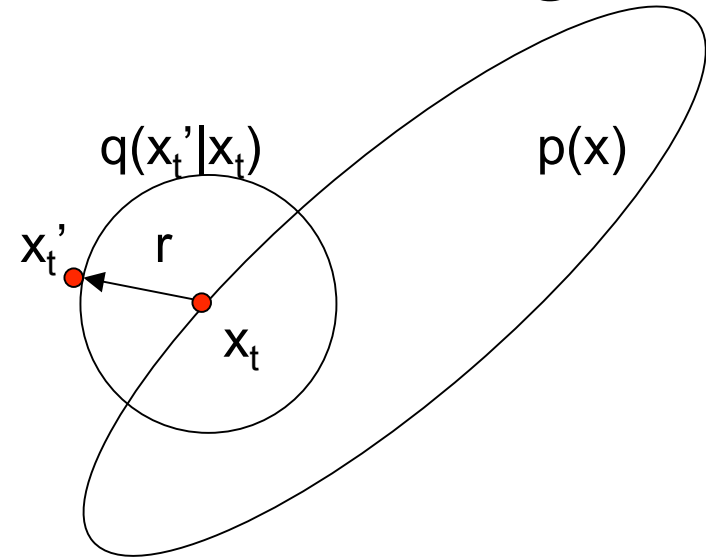
- I. Choose a candidate state x_t' using a *proposal distribution* $q(x_t'|x_t)$
- II. Compute the Metropolis ratio

$$r = \frac{p(x_t') p(d|x_t') q(x_t|x_t')}{p(x_t) p(d|x_t) q(x_t'|x_t)}$$

- III. If $r > 1$ then make the step: $x_{t+1} = x_t'$
if $r < 1$ then make the step with probability r , otherwise set $x_{t+1} = x_t$
so that the location is repeated
i.e., make the step with an *acceptance probability*

$$\alpha(x_t'|x_t) = \min \left\{ 1, \frac{p(x_t') p(d|x_t') q(x_t|x_t')}{p(x_t) p(d|x_t) q(x_t'|x_t)} \right\}$$

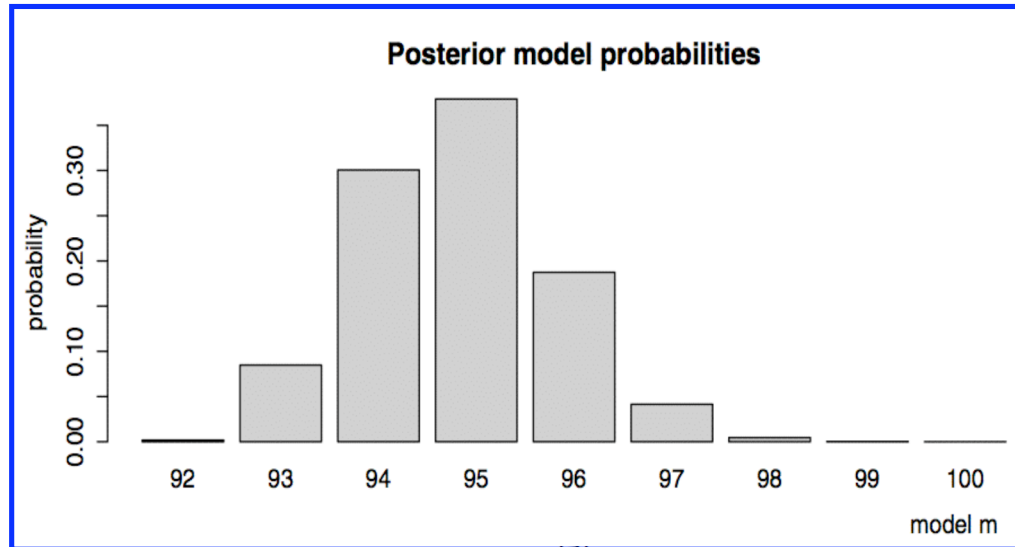
- IV. Choose next candidate based on the (new) current position...





Examples: Source confusion

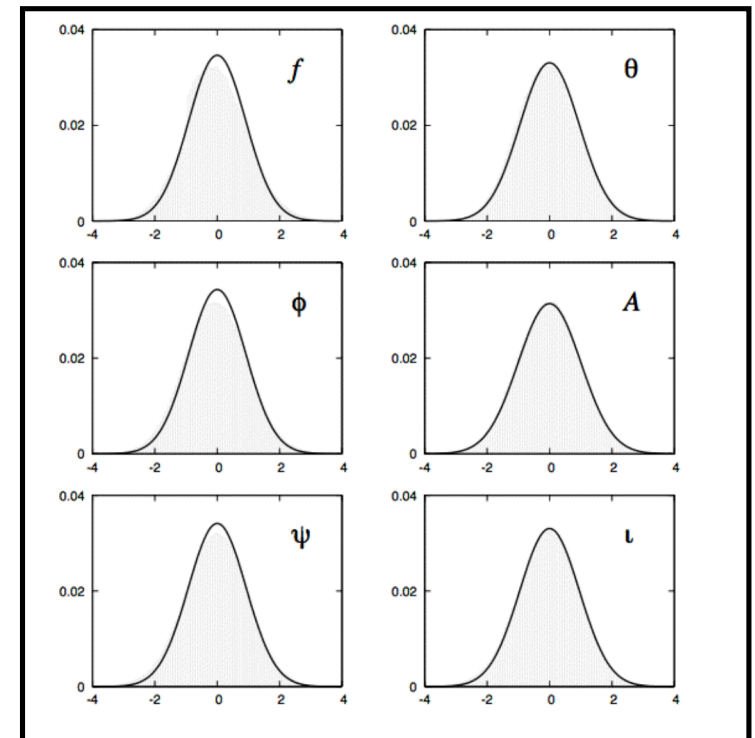
$N = 100$ sinusoids (N unknown)



(Umstaetter et al, 2005)

(Cornish and Crowder, 2005;
Cornish et al, 2006)

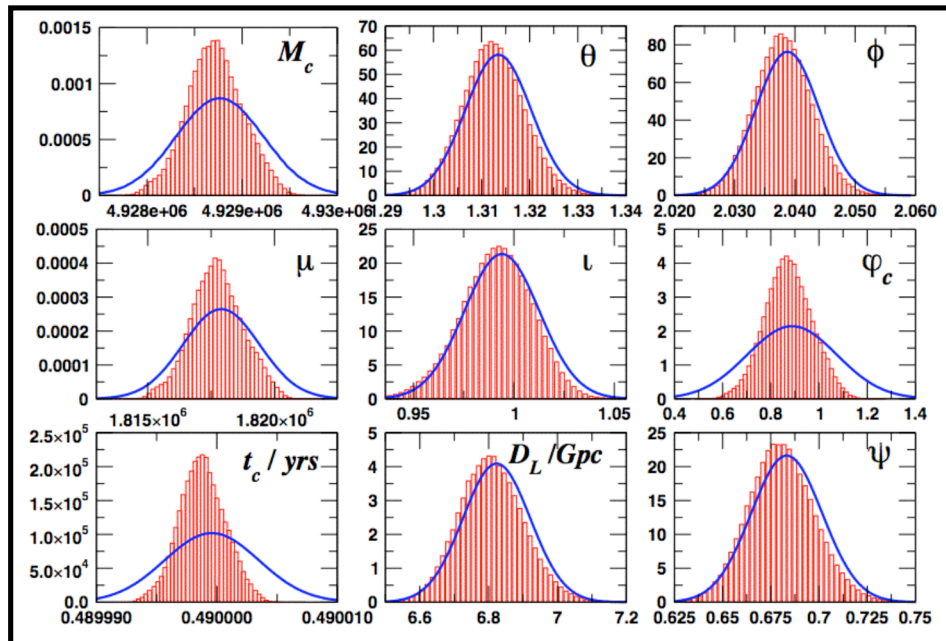
A few (N known) WD binaries





Example: MBHB (+ DWD)

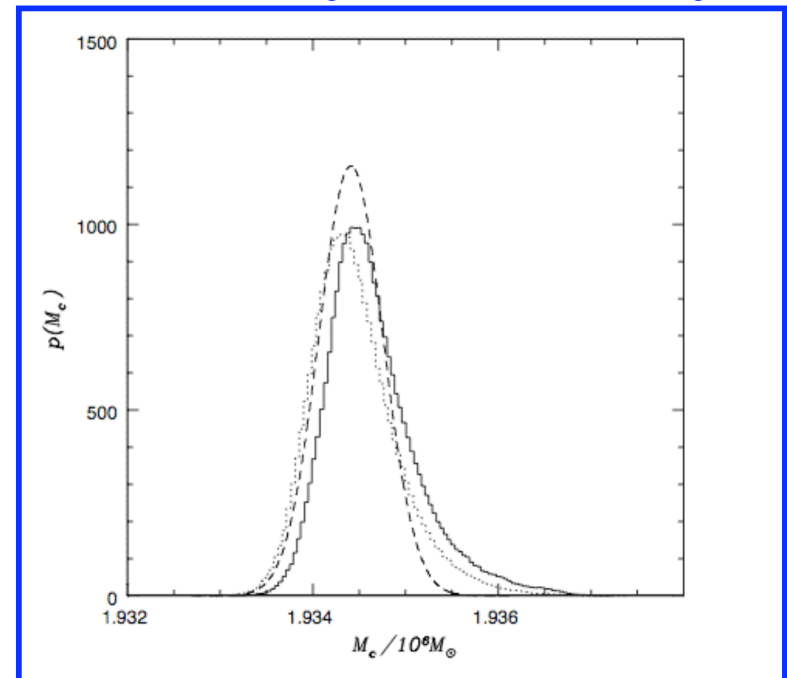
MBH binary



(Cornish and Porter, gr-qc/0605089

with foreground gr-qc/0605135)

MBH binary + WD binary



(Wickham, Stroeer, AV, gr-qc/0605071)



Matched filter

- Given the data set:

$$d(t) = h(t; \vec{\lambda}) + n(t)$$

- Construct a detection statistic c [here q is the [filter](#) or [template](#)]:

$$c = \int \tilde{d}(f) \tilde{q}^*(f) df$$

- The signal to noise ratio is:

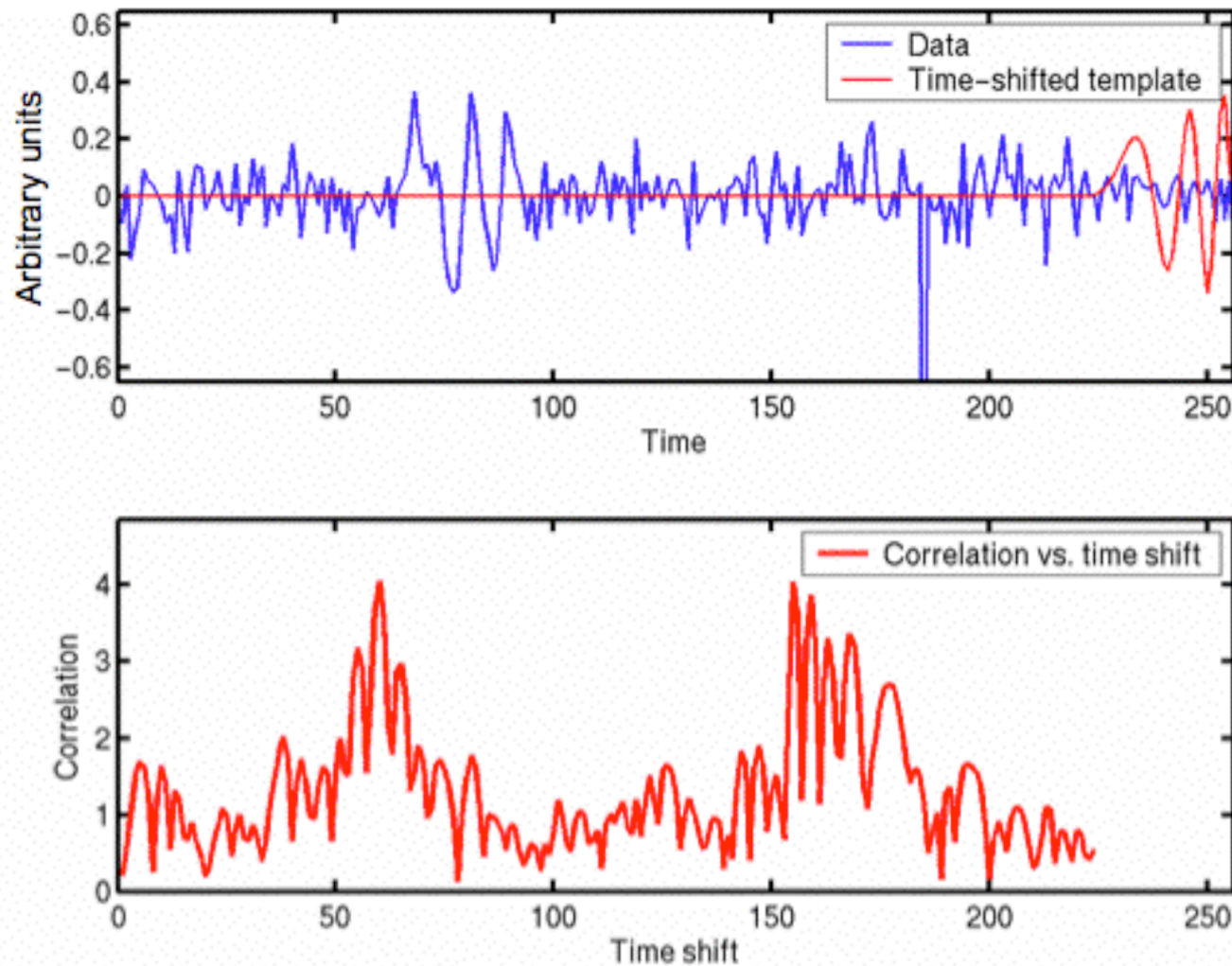
$$(S/N) = \frac{E[c]}{\sigma_c}$$

- Optimal filter (i.e. highest SNR) is:

$$\tilde{q}(f) \longrightarrow k \tilde{h}(f; \vec{\lambda}) / S(f)$$



Matched filter - an example





Signal-to-noise ratio (SNR)

- The (matched-filter) SNR or optimal SNR is:

$$(S/N)^2 = \langle h|h \rangle = 4 \int_0^\infty \frac{|\tilde{h}(f)|^2}{S(f)} df$$

- The optimal SNR scales as the sqrt of the integration time (or the number of recorded wave cycles): **one can “dig into” the noise:**

$$(S/N)^2 \propto \int \mathcal{N}(f) \frac{h^2[t(f)]}{h_{\text{rms}}^2(f)} d(\ln f)$$

$$\propto f_c T \frac{h^2}{h_{\text{rms}}^2} \quad \leftarrow h \text{ can be } \ll h_{\text{rms}}$$

$$10^4 (f/1 \text{ mHz}) (T/10^7 \text{ s})$$



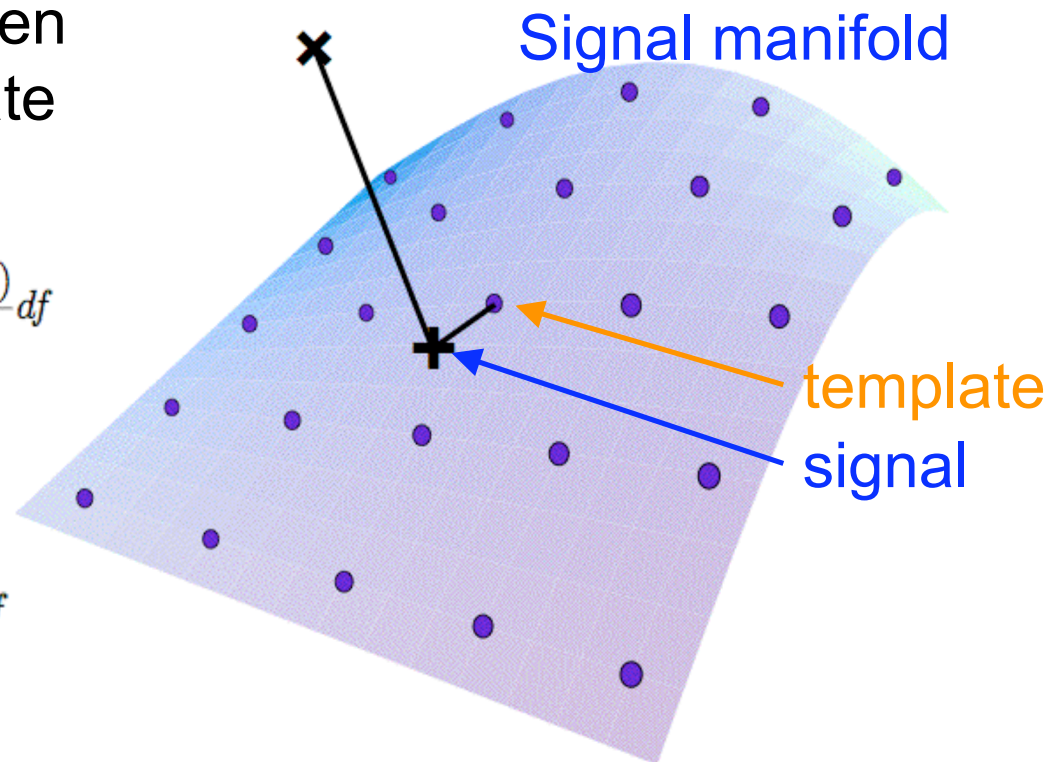
Geometric approach to data analysis

- Scalar product between a signal and a template (or two signals):

$$\langle g|h \rangle = 2 \int_0^\infty \frac{\tilde{g}^*(f)\tilde{h}(f) + \tilde{g}(f)\tilde{h}^*(f)}{S(f)} df$$

- Signal-to-noise ratio:

$$(S/N)^2 = \langle h|h \rangle = 4 \int_0^\infty \frac{|\tilde{h}(f)|^2}{S(f)} df$$



Dhurandhar and Sathyaprakash (1993)
Owen (1996)



Fisher information matrix

- Likelihood (for $S/N \gg 1$):

$$p(\vec{\lambda}|s) \propto \exp\left[-\frac{1}{2}\Gamma_{ab}\Delta\lambda^a\Delta\lambda^b\right]$$

- The Fisher information matrix is: $\Gamma_{ab} = \langle \partial_a h | \partial_b h \rangle$
- The variance-covariance matrix Σ is the inverse of Γ
- Statistical mean square errors and correlation coefficients:

$$\sigma_a = \langle (\Delta\lambda^a)^2 \rangle^{1/2} = \sqrt{\Sigma^{aa}} \quad c^{ab} = \frac{\langle \Delta\lambda^a \Delta\lambda^b \rangle}{\sigma_a \sigma_b}$$

- Note: this provides a **lower bound** to the error (Cramer-Rao bound)

Parameter determination: MBHB

MBH binary systems

$$m_1 = 10^6 M_{\text{sun}}$$

$$m_2 = 10^6 M_{\text{sun}}$$

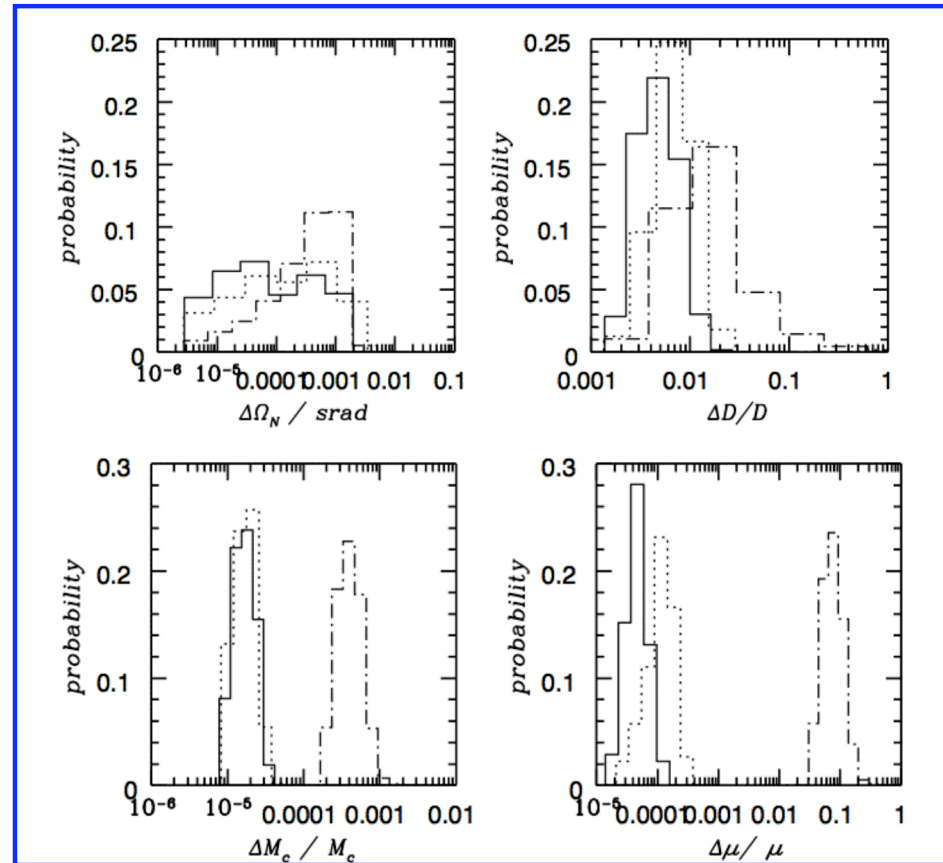
$$z = 1$$

- Error box in the sky $\sim 1 \text{ deg}^2$

$$\forall \Delta D/D \sim 0.01 - 0.1$$

$$\forall \Delta m/m \sim 10^{-4} - 0.1$$

*Many papers: Cutler (1998),
Hughes, Holz, Cornish, Krolak,
Buonanno, Berti,
Will, Sathyaprakash, AV...*



(AV, 2004)

Parameter determination: EMRI

$$\frac{\Delta m}{m} \sim \frac{\Delta M}{M} \sim \Delta\left(\frac{S}{M^2}\right) \sim 10^{-4} \quad \Delta\theta \sim 2^\circ \quad \frac{\Delta D}{D} \sim 0.05$$

S/M^2 e_{LSO}	0.1 0.1	0.1 0.3	0.1 0.5	0.5 0.1	0.5 0.3	0.5 0.5	1 0.1	1 0.3	1 0.5
$\Delta(\ln M)$	2.6e-4	5.6e-4	5.3e-5	2.7e-4	9.2e-4	7.7e-5	2.8e-4	2.5e-4	1.5e-4
$\Delta(S/M^2)$	3.6e-5	7.9e-5	4.5e-5	1.3e-4	6.3e-4	5.1e-5	2.6e-4	3.7e-4	2.6e-4
$\Delta(\ln \mu)$	6.8e-5	1.5e-4	7.4e-5	6.8e-5	9.2e-5	1.0e-4	6.1e-5	9.1e-5	1.0e-3
$\Delta(e_0)$	6.3e-5	1.3e-4	2.9e-5	8.5e-5	2.8e-4	3.2e-5	1.2e-4	1.1e-4	1.6e-4
$\Delta(\cos \lambda)$	6.0e-3	1.7e-2	1.3e-3	1.3e-3	5.8e-3	2.4e-4	6.5e-4	8.4e-4	4.7e-4
$\Delta(\Omega_s)$	1.8e-3	1.7e-3	7.9e-4	2.0e-3	1.7e-3	7.6e-4	2.1e-3	1.1e-3	6.7e-4
$\Delta(\Omega_K)$	5.6e-2	5.3e-2	4.7e-2	5.5e-2	5.1e-2	4.7e-2	5.6e-2	5.1e-2	4.8e-2
$\Delta(\bar{\gamma}_0)$	4.0e-1	6.3e-1	3.8e-1	1.0e+0	6.1e-1	3.9e-1	9.3e-1	3.4e-1	3.9e-1
$\Delta(\Phi_0)$	2.6e-1	6.7e-1	2.2e-1	1.4e+0	7.5e-1	2.7e-1	1.5e+0	1.7e-1	3.3e-1
$\Delta(\alpha_0)$	6.2e-1	5.8e-1	5.5e-1	6.3e-1	5.9e-1	5.6e-1	6.4e-1	5.9e-1	5.9e-1
$\Delta[\ln(\mu/D)]$	8.7e-2	3.8e-2	3.7e-2	3.8e-2	3.7e-2	3.7e-2	3.8e-2	7.0e-2	3.7e-2
$\Delta(t_0)\nu_0$	4.5e-2	1.1e-1	3.3e-2	2.3e-1	1.3e-1	4.4e-2	2.5e-1	3.2e-2	5.5e-2

(For $10 M_\square$ onto $10^6 M_\square$ at 1Gpc, for various eccentricities and spins)

Barack and Cutler (2004)



Geometric approach to data analysis

Match:

$$M(\vec{\lambda}, \Delta\vec{\lambda}) \equiv \max_{\mu, \Delta\mu} \langle u(\vec{\mu}, \vec{\lambda}) | u(\vec{\mu} + \Delta\vec{\mu}, \vec{\lambda} + \Delta\vec{\lambda}) \rangle$$

$$M(\vec{\lambda}, \Delta\vec{\lambda}) \approx 1 + \frac{1}{2} \left(\frac{\partial^2 M}{\partial \Delta\lambda^i \partial \Delta\lambda^j} \right)_{\Delta\lambda^k=0} \Delta\lambda^i \Delta\lambda^j$$

$$g_{ij}(\vec{\lambda}) = -\frac{1}{2} \left(\frac{\partial^2 M}{\partial \Delta\lambda^i \partial \Delta\lambda^j} \right)_{\Delta\lambda^k=0}$$

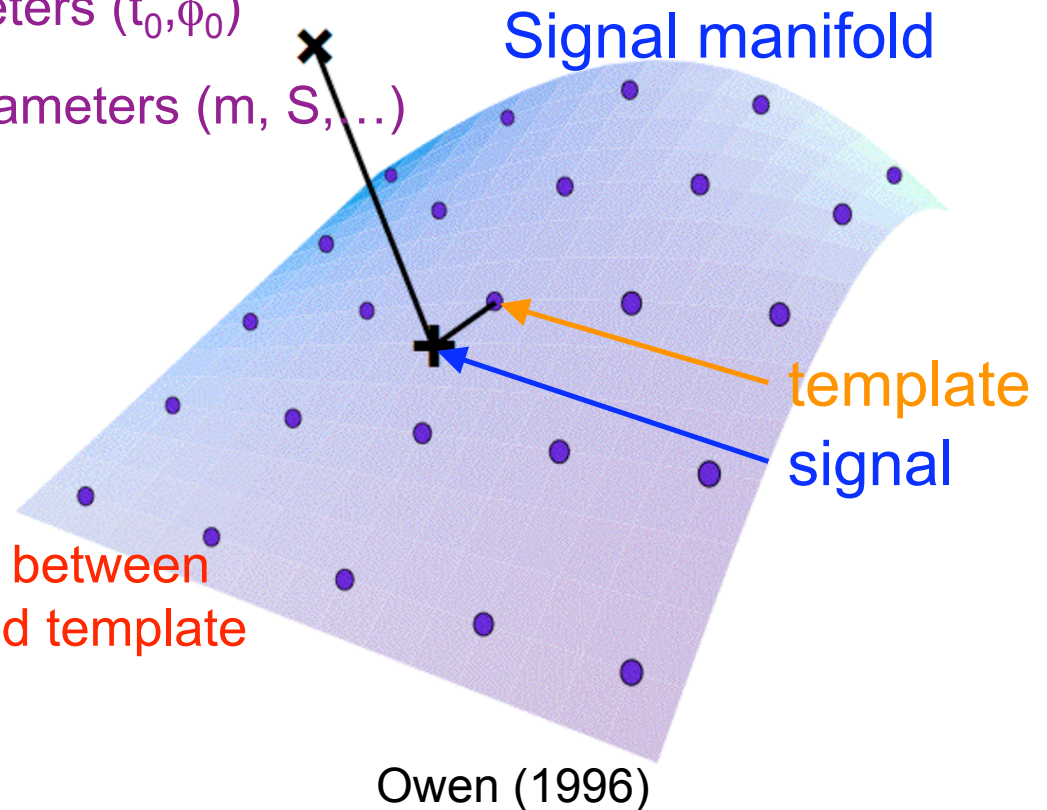
$$1 - M = g_{ij} \Delta\lambda^i \Delta\lambda^j$$

Distance between signal and template

Metric on parameter space

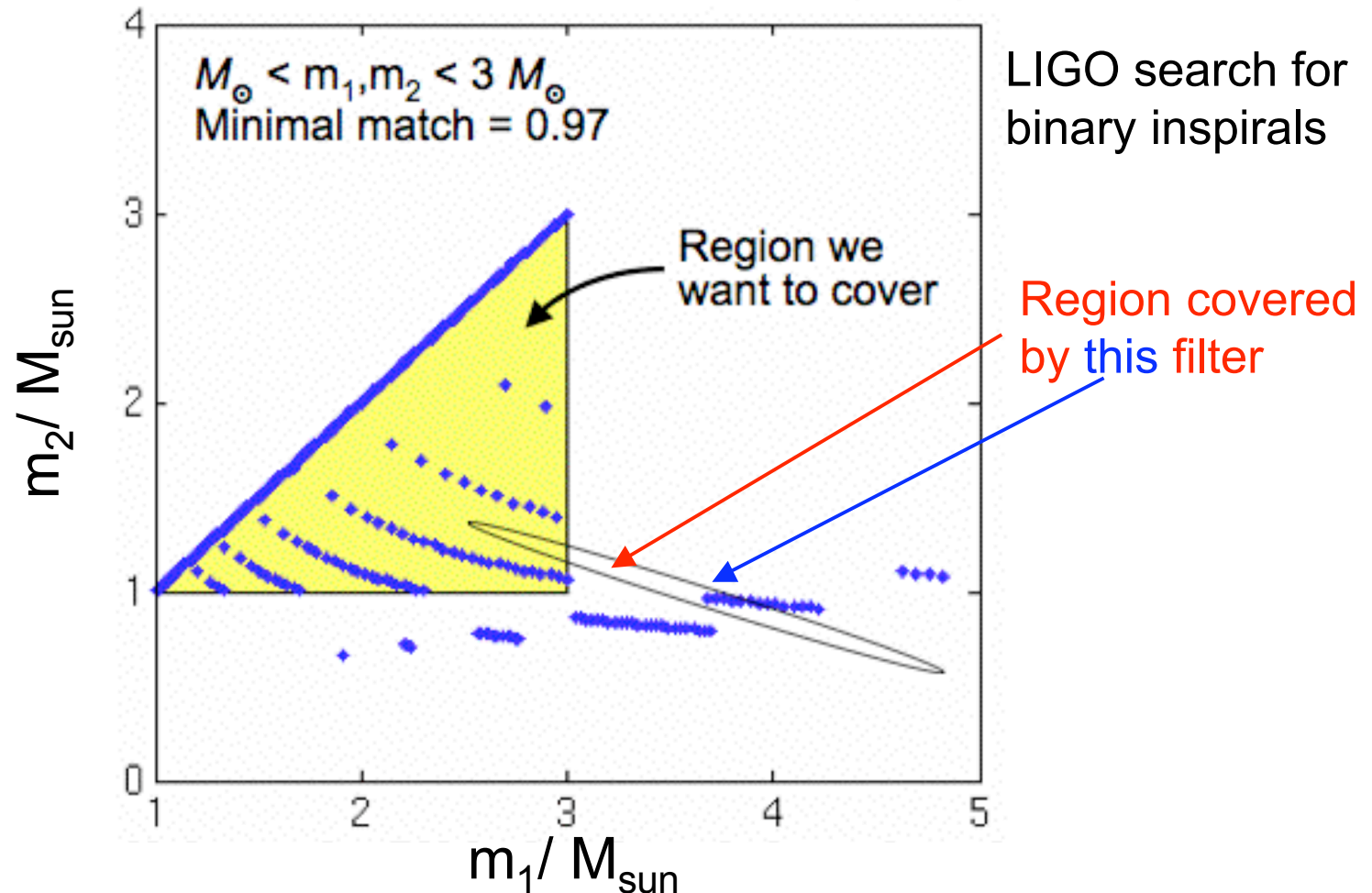
Extrinsic parameters (t_0, ϕ_0)

Intrinsic parameters (m, S, \dots)



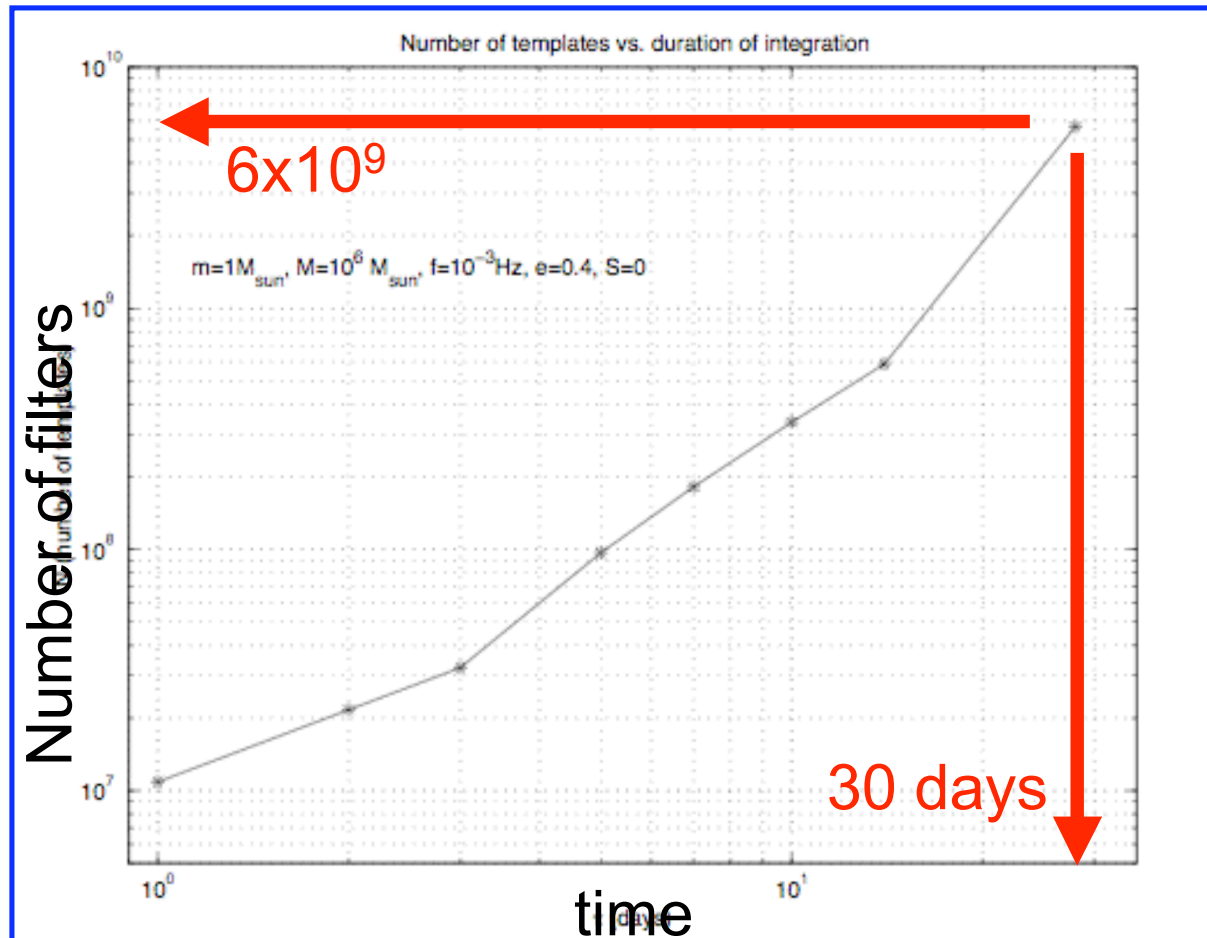


Example: template bank for binary in-spiral





Computational costs for EMRI



(Barack and Cutler)

19th June 2006

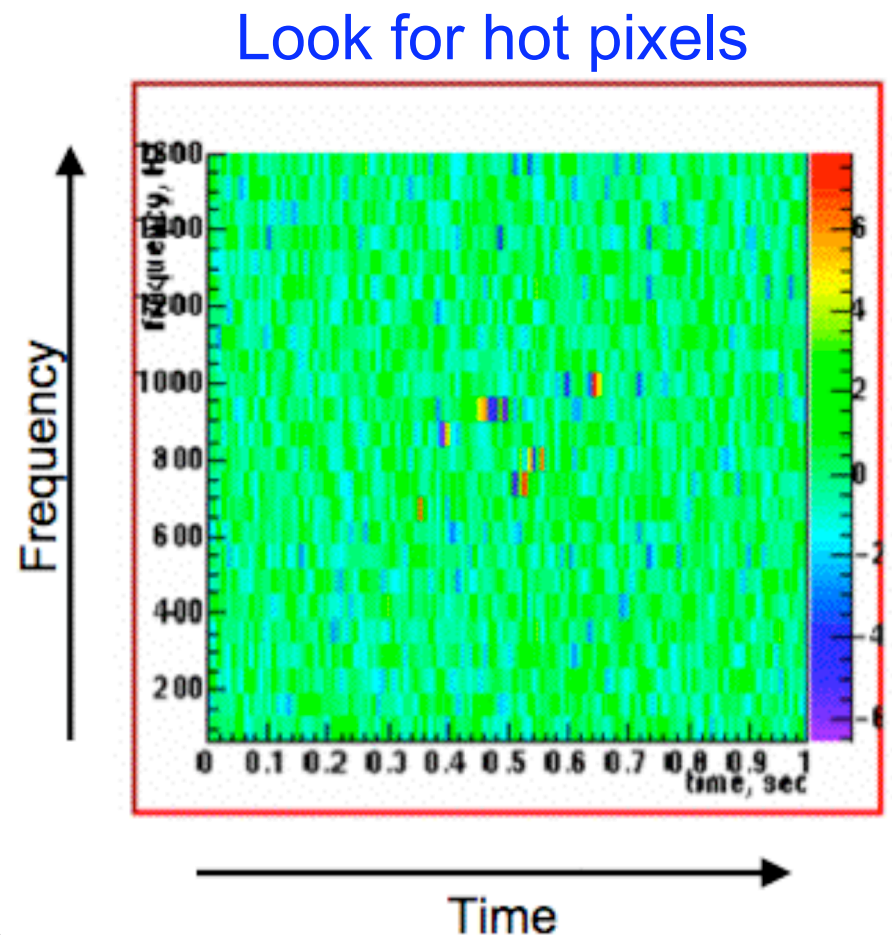
A Vecchio - LISA Data Analysis

34



Sub-optimal methods

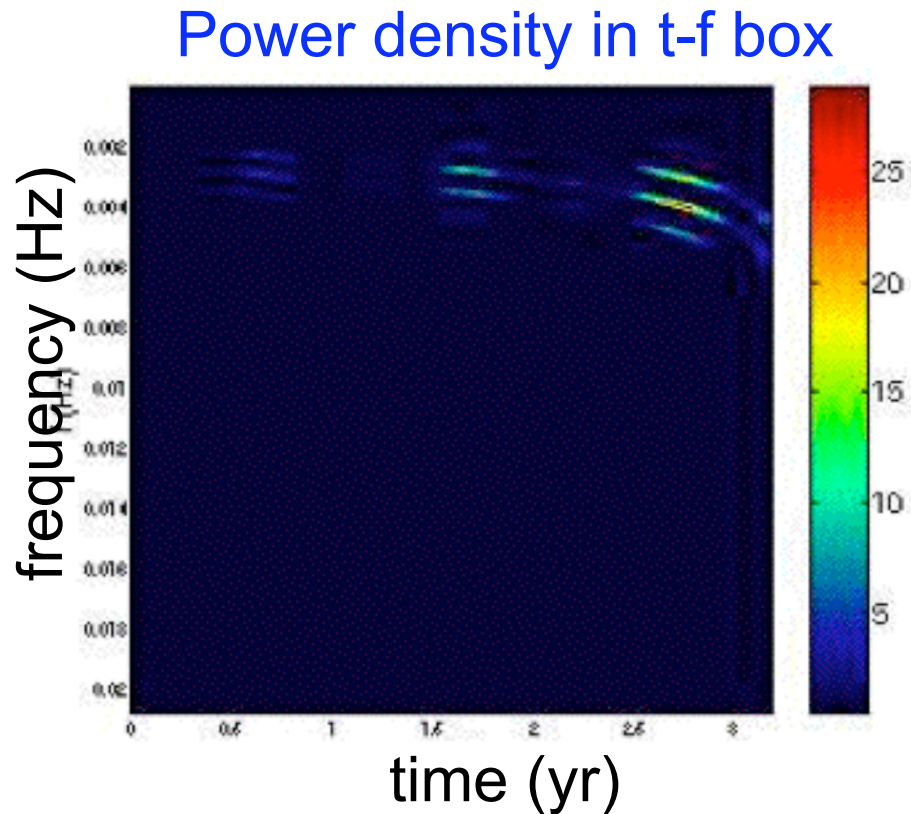
- Matched filtering is not a viable strategy if:
 - Theoretical waveforms are not accurate enough (such as poorly modeled burst signals - e.g. final plunge of MBH binary)
 - Computational costs are too high (e.g. EMRI)
- Alternatives:
 - “Incoherent methods”
 - Hierarchical methods



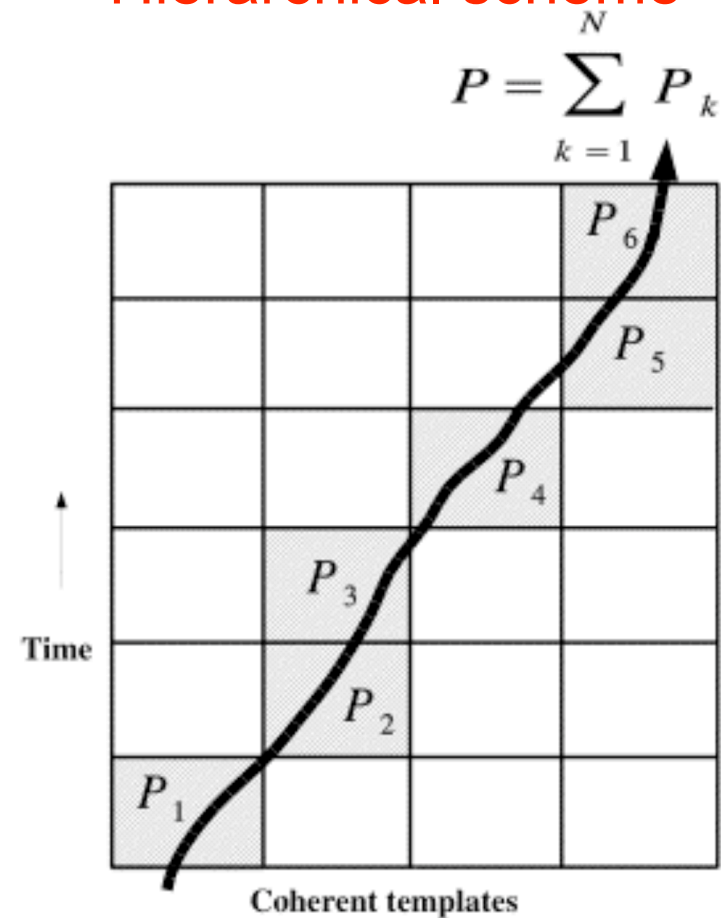


An example: EMRI

Hierarchical scheme



(Wen and Gair, 2005)

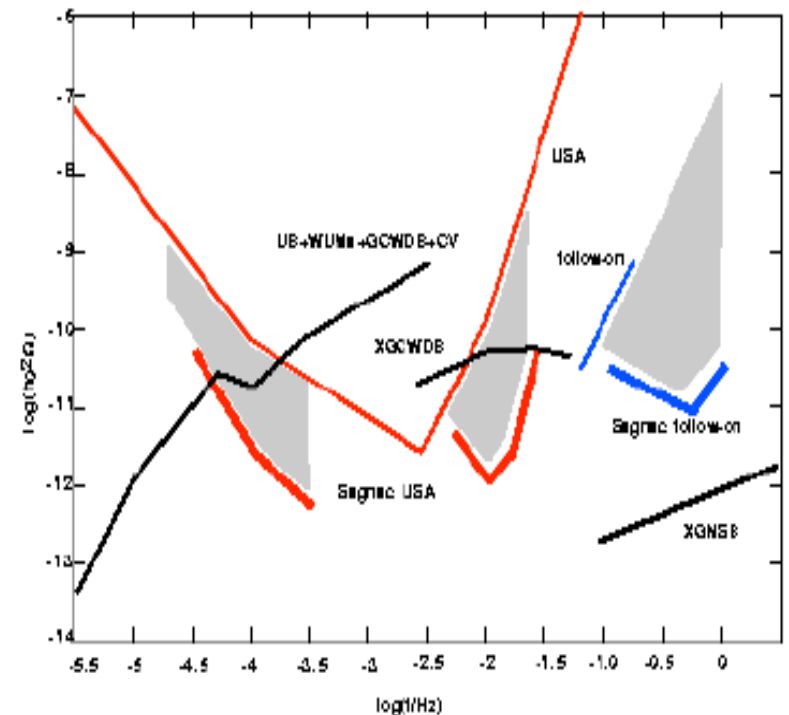


(T Creighton, Gair, ..)



Stochastic signals

- The detection of stochastic signals depends on LISA multiple observable
 - Cross-correlation (a la LIGO) are not possible
 - Symmetrized-Sagnac (essentially insensitive to GWs at low frequency) is likely the key
- As of today we do not have a solid strategy to detect an isotropic stochastic signal



(Hogan and Bender, 2001)



Conclusions

- LISA data analysis poses a wide spectrum of interesting problems:
 - Specific to LISA and GW observations
 - General with implications for other fields of astronomy
- The outstanding issues are gradually being resolved
- ... but there is still a lot of work to be done
- Many more details in plenary and parallel sessions



A new way to look at TDI

- Consider 2 data streams:

$$s_1 = p + n_1 + h_1$$

$$s_2 = p + n_2 + h_2$$

noises:

p is common noise: $\langle p \rangle = 0$ and $\langle p^2 \rangle = \sigma_p^2$

$n_{1,2}$: $\langle n_1^2 \rangle = \langle n_2^2 \rangle = \sigma_n^2$

n and p are uncorrelated: $\langle n_1 n_2 \rangle = \langle n_1 p \rangle = \langle n_2 p \rangle = 0$



All info are in the likelihood

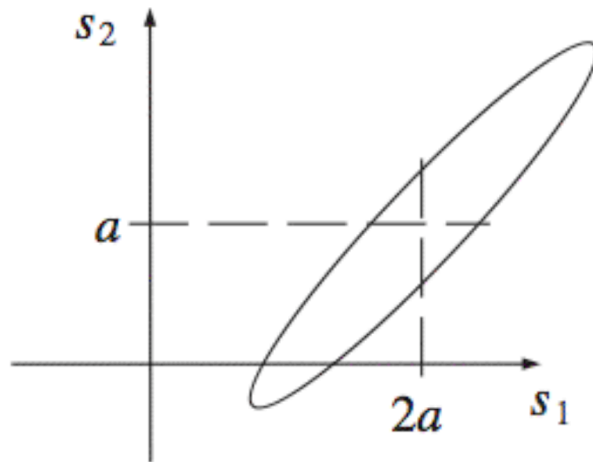
- Likelihood:

$$p(s_1, s_2|a) \propto \exp\left[-\frac{1}{2}Q\right]$$

$$Q \equiv (\mathbf{s} - \mathbf{h})^T \cdot C^{-1} \cdot (\mathbf{s} - \mathbf{h})$$

$$\equiv \sum_{i,j=1}^2 (s_i - h_i) C_{ij}^{-1} (s_j - h_j)$$

- C is noise covariance matrix



$$C_{ij} \equiv \langle (s_i - h_i)(s_j - h_j) \rangle$$

$$C = \begin{pmatrix} \sigma_p^2 + \sigma_n^2 & \sigma_p^2 \\ \sigma_p^2 & \sigma_p^2 + \sigma_n^2 \end{pmatrix}$$



Principal component analysis

- Find eigenvalues/vector of C and diagonalize:

$$p(s_1, s_2|a) \propto p(s_+|a)p(s_-|a),$$

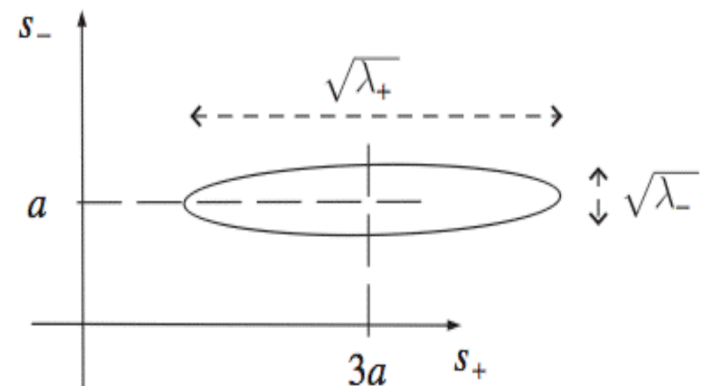
$$s_+ \equiv s_1 + s_2,$$

$$s_- \equiv s_1 - s_2.$$

$$p(s_+|a) \propto \exp\left[-\frac{1}{2} \frac{(s_+ - 3a)^2}{4\sigma_p^2 + 2\sigma_n^2}\right],$$

$$p(s_-|a) \propto \exp\left[-\frac{1}{2} \frac{(s_- - a)^2}{2\sigma_n^2}\right].$$

- For LISA $\sigma_p^2 \gg \sigma_n^2$, so there is no loss statistical inference only on the s_- term observable (Romano and Woan, 2006)





Noise

- Assuming Gaussian and stationary noise:

- Mean: $\langle \tilde{n}(f) \rangle = 0$

- Variance $\langle \tilde{n}(f) \tilde{n}^*(f') \rangle = \frac{1}{2} S(f) \delta(f - f')$

- rms fluctuation of the noise in a band Δf is:

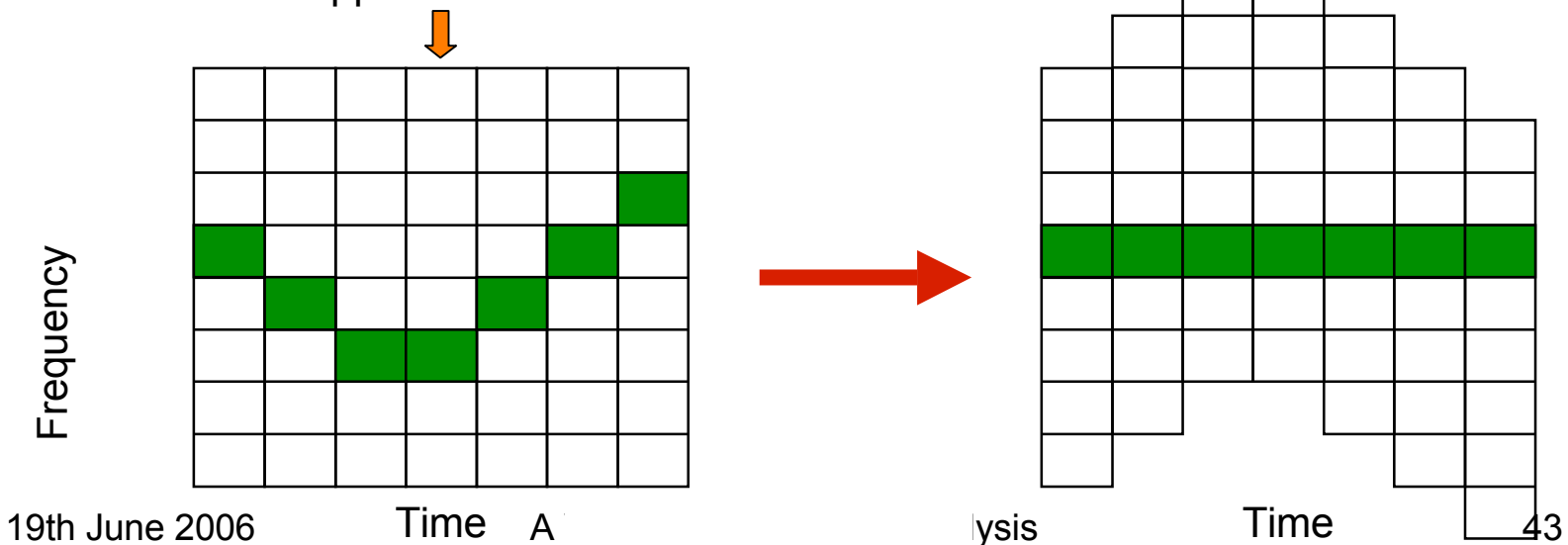
$$h_{\text{rms}}(f) = \sqrt{\Delta f S(f)}$$

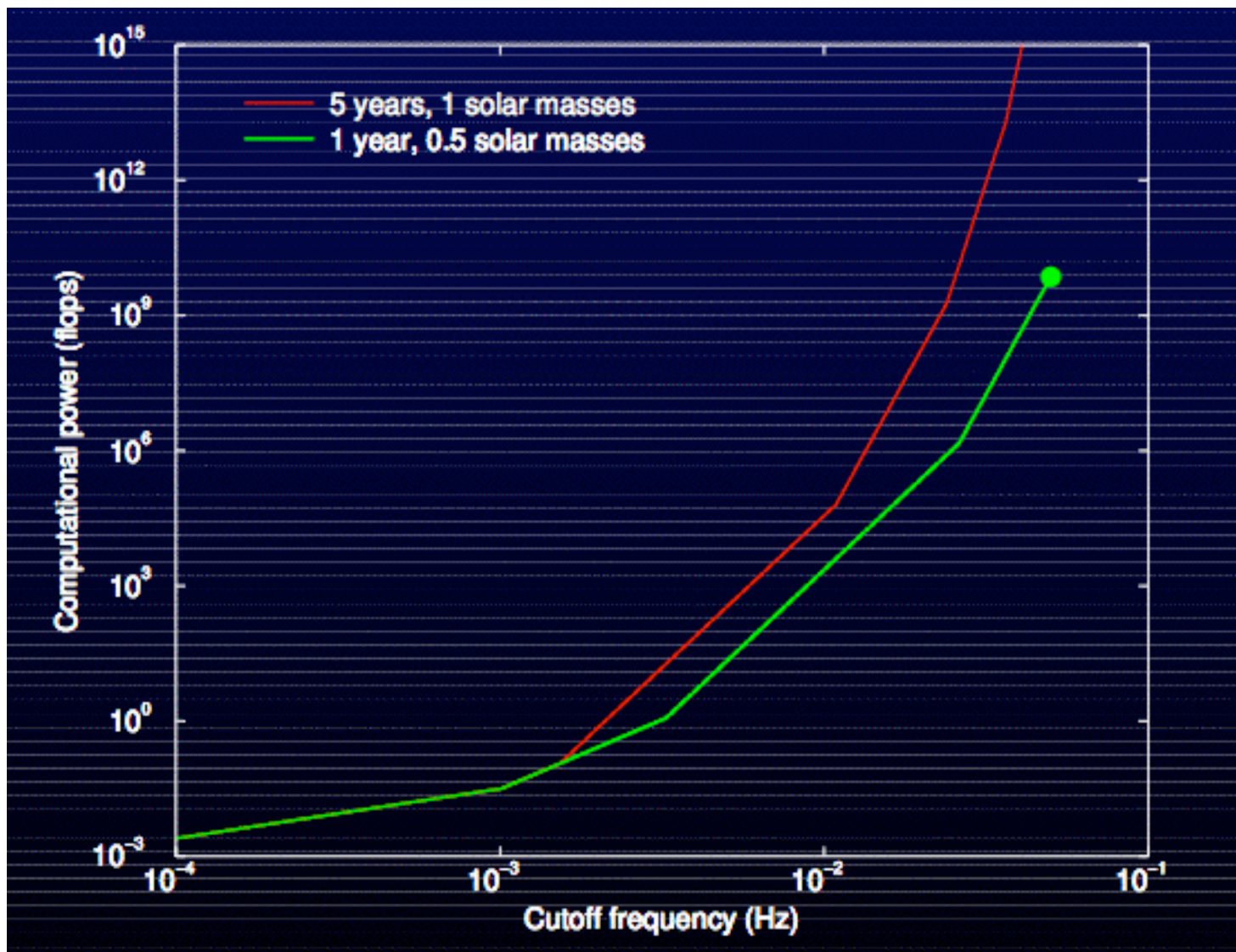
$$h_{\text{rms}}(f) = \sqrt{f S(f)}$$

The StackSlide Method

- Break up data into segments; FFT each, producing Short (30 min) Fourier Transforms (SFTs) = coherent step.
- StackSlide: stack SFTs, track frequency, slide to line up & add the power weighted by noise inverse = incoherent step.
- Other semi-coherent methods:
 - Hough Transform: Phys. Rev. D72 (2005) 102004; gr-qc/0508065.
 - PowerFlux: see next talk, W11.00005
- Fully coherent methods:
 - Frequency domain match filtering/maximum likelihood estimation (C7.00001; W11.00006)
 - Time domain Bayesian parameter estimation (C7.00002)
- **Improvements and hierarchical pipeline under development.**

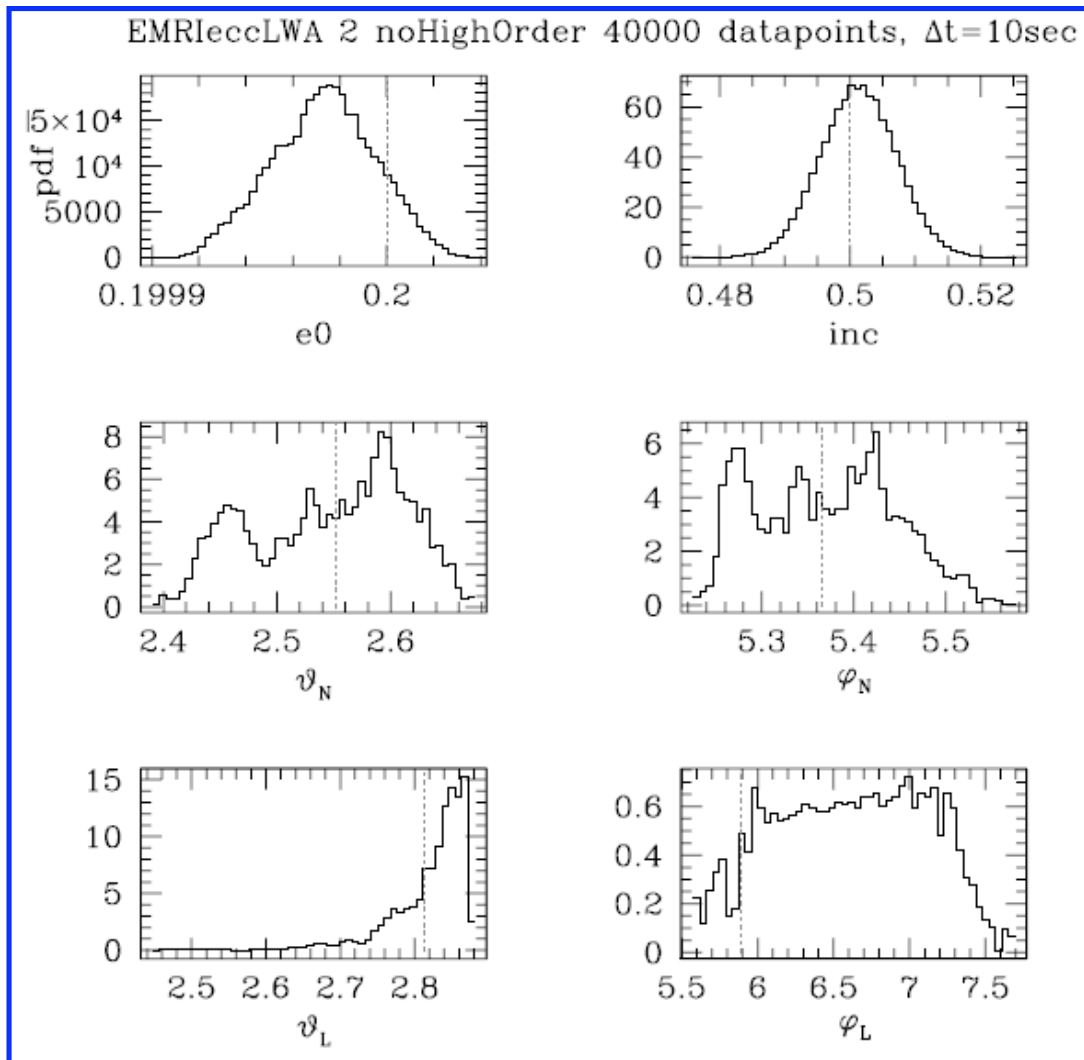
Track Doppler shift and df/dt







Example: EMRI



(Stroeer & Gair)